

牛莉霞, 赵月, 李国, 等. 人机信任对矿工不安全行为的影响: 基于认知-情感双路径机制[J]. 安全与环境学报, 2026, 26(6): 2242-2252.

NIU L X, ZHAO Y, LI G, et al. Exploring the influence mechanism of human-machine trust on miners' unsafe behaviors: a cognitive-emotional dual-path approach[J]. Journal of Safety and Environment, 2026, 26(6): 2242-2252.

人机信任对矿工不安全行为的影响: 基于认知-情感双路径机制*

牛莉霞¹, 赵月¹, 李国², 李肖萌¹

(1 辽宁工程技术大学工商管理学院, 辽宁葫芦岛 125105; 2 中矿检测(辽宁)有限公司, 辽宁阜新 123000)

摘要:为系统探究智能化煤矿背景下人机信任对矿工不安全行为的作用机制, 基于认知-情感个性系统理论与社会认知理论, 构建以风险感知和技术焦虑为中介、任务复杂性为调节变量的双路径影响模型。通过对山西省、辽宁省、内蒙古自治区 508 名一线矿工开展问卷调查, 运用结构方程模型与层次回归分析法对相关假设进行实证检验。结果显示: 人机信任通过削弱风险感知显著增加不安全行为, 同时, 经缓解技术焦虑间接抑制不安全行为, 呈现出“双刃剑”效应。进一步分析发现, 任务复杂性正向调节人机信任对风险感知的影响, 负向调节人机信任对技术焦虑的影响, 且在风险感知路径中起正向调节作用, 在技术焦虑路径中起负向调节作用。研究丰富了智能化矿山安全行为领域的理论体系, 为提高矿工安全行为和完善智能矿山人机协作管理提供了理论参考与实践依据。

关键词:安全社会科学; 人机信任; 矿工不安全行为; 认知-情感双路径; 风险感知; 技术焦虑; 任务复杂性

中图分类号: X915.2

文献标志码: A

文章编号: 1009-6094(2026)06-2242-11

DOI: 10.13637/j.issn.1009-6094.2025.1066

0 引言

煤矿产业是我国能源安全的支柱, 安全生产一直备受重视^[1]。近年来, 随着人工智能和自动化技术的引入, 煤矿企业重塑了传统的人机交互模式^[2], 矿工的不安全行为已从操作失误转向由信任偏差引起的认知与情感失衡。智能化系统通过高效的感知和预警能力提升了矿井的安全管理水平^[3], 但矿工对智能系统不稳定的信任可能导致心理和生理失衡, 从而产生误操作并引发事故^[4]。当矿工对智能化系统信任过高时, 易产生过度依赖, 忽视自身判断和安全规程, 增加不安全行为的发生概率; 而当信任不足时, 则可能忽视系统预警, 甚至抗拒其建议, 采取违规操作以弥补感知的不确定性。因此, 信任与实际系统可靠性不匹配, 会影响矿工的安全决策和行为表现。如何在智能系统普及背景下理解和干预矿工不安全行为, 成为安全管理亟待破解的现实难题。

随着煤矿智能化技术的快速发展, 人机信任成为矿工与智能系统交互中的重要心理因素, 逐渐影响矿工的行为决策。现有研究主要聚焦于矿工个体^[5]、环境^[6]及领导^[7]等传统因素, 鲜有学者探讨

人机信任如何作用于不安全行为。1987 年, Muir^[8]首次提出人机信任的概念, 即用户对机器能力、可靠性和意图的积极预期。由于人机信任来源于矿工与智能系统的长期交互, 因此能够影响矿工对系统的依赖程度和风险判断。当前, 矿工的风险意识和应急经验相对薄弱, 容易在高自动化场景中产生过度信任或盲目信赖, 从而引发认知偏差和不安全行为。适度的人机信任有助于提升矿工对系统操作的信心, 缓解技术焦虑情绪, 促进安全行为。因此, 探究人机信任对矿工不安全行为的影响路径, 识别其中的认知与情感机制, 对于促进矿工安全行为和减少事故风险具有重要意义。

最初, 人机信任研究集中于用户对技术系统能力、可靠性和意图的预期, 广泛应用于航空、军事等高自动化领域。随着 AI 和自动化系统的普及, 研究者开始关注人机信任对行为偏差的影响。Wiczorek 等^[9]指出人机信任过度易引发自动化偏差, 操作人员可能忽视系统错误的建议, 增加安全风险。Hancock 等^[10]强调信任通过反馈与经验持续调整能减少依赖偏差并增强系统的风险监控能力。Chien 等^[11]发现, 驾驶员对自动驾驶系统的高度信任削弱了其风险感知和警觉性, 从而降低了应对突发情况

* 收稿日期: 2025-06-19

作者简介: 牛莉霞, 教授, 博士, 从事组织行为管理与人因工程方面的研究, 327334231@qq.com。

基金项目: 国家自然科学基金项目(52174184)

的能力。尽管人机信任相关研究不断拓展应用边界,但在智能化矿山环境中,关于其对矿工不安全行为影响的研究仍较少。因此,亟须进一步探索人机信任通过认知与情感路径对矿工不安全行为的影响机制,该研究视角不仅在理论上具有补充意义,也可以为提升矿井安全管理效能与作业人员行为干预策略提供实践参考。

鉴于此,本文基于认知-情感个性系统理论与社会认知理论,探讨人机信任如何通过认知与情感双路径影响不安全行为,并分析风险感知和技术焦虑的中介作用,以及任务复杂性的调节作用,旨在为智能化矿山的安全管理优化、矿工行为干预和安全素养提升提供理论指导和实践依据。

1 理论基础与研究假设

1.1 人机信任与不安全行为

人机信任是个体在不确定或易受伤害的情境下,认为代理能助其达到目标的态度^[12]。近年来,信任机制与不安全行为关联的研究备受关注:情感层面的信任可影响作业人员的不安全行为^[13],矿工群体内部信任关系强度与不安全行为抑制效果呈正相关^[14]。这表明人际信任与不安全行为联系紧密,而人机信任研究起源于人际信任^[15],不安全行为是矿工作业中导致事故的关键人为因素。在已经习惯于与人类合作的智能煤矿领域中,人们越来越多地将原先的互动对象从人类替换到AI,因此,人际信任可以作为人机信任研究的基础。

社会认知理论指出,个体的心理与情感因素可通过影响其认知过程进而作用于行为^[16]。人机信任改变矿工的认知,进而显著影响其行为。与系统可靠性不匹配的人机信任可能会对矿工行为产生负面影响,人机信任通过与智能系统的长期交互逐步建立,与人际信任不同,其主要影响矿工对系统的依赖程度与风险判断能力。当信任缺失时,工人群体的不安全行为会受到抑制,传播速度减缓^[17]。良好的人机信任提升矿工的安全感、归属感和心理稳定性,进而促进遵守安全规范并减少不安全行为。员工间信任能促进相互关注与纠正不安全行为,抑制其扩散^[14]。在信任氛围下,员工更愿意尝试新技术,倾向于持续采纳人工智能技术^[18],减少技术焦虑,降低不安全行为的发生概率。综上,人机信任对矿工不安全行为的影响具有“双刃剑”效应,其影响机制需通过认知和情感双路径深入剖析。

1.2 风险感知的中介作用

风险感知是个体基于经验和主观感受对客观风

险的认知^[19]。根据认知-情感个性系统理论,个体通过理性与情感的结合,依据外部情景激发的认知来决定是否采取特定行为^[20]。在智能化煤矿场景中,人机信任成为影响风险判断的重要心理前提。当作业人员对AI系统的信任水平较高时,更可能认为系统可靠、安全,从而低估操作环境中的潜在风险,即风险感知水平降低^[21]。反之,若对系统缺乏信任,则更倾向于认为AI存在不确定性,从而高估作业风险,提高风险感知水平。

风险感知则直接影响行为决策。当矿工的风险感知水平较高时,能够更准确地识别潜在危险并采取防范措施,降低不安全行为发生概率;当风险感知不足时,则容易忽视风险征兆,导致不当操作,增加不安全行为的发生概率^[22]。人机信任对风险感知具有负向影响,风险感知对不安全行为具有负向影响,因此风险感知在人机信任与不安全行为之间构成正向的间接关系。据此,提出以下假设

H₁:风险感知在人机信任与不安全行为之间起中介作用,使得人机信任水平与不安全行为之间存在正向的间接关系。

1.3 技术焦虑的中介作用

技术焦虑源于先进技术副作用,体现为非理性恐惧或焦虑情绪^[23]。根据认知-情感个性系统理论,个体情感单元可被外部情景激发,这些被激活的情感反应是影响行为的关键^[20]。AI系统的复杂性与不透明性可能引发员工对技术的不安与焦虑。人机信任是影响技术焦虑的重要前置变量。当个体对AI系统的可靠性、安全性和可控性具有较高信任时,更容易形成积极的心理预期,从而减缓不确定性带来的情绪冲击,降低技术焦虑水平^[24];反之,缺乏信任则增强技术焦虑。

积极情绪有助于员工执行标准化安全流程,而焦虑等负面情绪则可能造成注意力分散、决策偏差与行为失控,进而提高不安全行为发生的概率^[25]。人机信任通过负向影响技术焦虑,影响作业人员的情绪与行为反应,进而对不安全行为产生间接的负向影响。据此,提出以下假设

H₂:技术焦虑在人机信任与不安全行为之间起中介作用,使得人机信任和不安全行为之间有负向的间接关系。

1.4 任务复杂性的调节作用

任务复杂性包括认知负荷、信息处理需求和操作难度。根据社会认知理论,个体因素通过影响认知过程作用于行为^[16]。在矿山智能作业环境中,作

业人员在高风险、高强度的任务中,任务复杂性显著影响其对环境信息的处理能力。当任务复杂性较高时,操作人员通常面临信息超载和时间压力,此时更容易依赖 AI 系统进行判断与决策。当矿工对 AI 系统信任时,这种依赖倾向尤为明显,可能导致个体忽视自身的风险评估与经验判断,从而降低对潜在风险的敏感度,风险感知水平随之下降。个体对 AI 系统的信任可能会削弱其对系统潜在失败或不确定性的警觉性^[26]。个体为完成复杂任务而优先调动有限的注意资源,进一步减少对潜在风险的关注,从而放大大机信任对风险感知的负向作用^[27]。相反,在任务复杂性较低时,操作人员的认知资源相对充足,更有能力整合自身经验与 AI 建议,独立进行风险评估,从而削弱人机信任对风险感知的消极影响。据此,提出以下假设

H₃:任务复杂性正向调节人机信任和风险感知之间的关系。对于任务复杂性较高的作业,人机信任对风险感知的负向影响更强。

根据社会认知理论,个体认知与情绪状态受环境条件与个体信念的共同作用^[16]。人机信任可通过增强个体对 AI 系统的可预测性与可靠性感知,缓解技术带来的不确定感与焦虑情绪^[24]。复杂任务能提供更具有挑战性与价值感的工作体验,进而增强员工的内在动机^[28],因此任务复杂性水平的变化,可能影响这一缓解效应的强度。在高复杂性任务的情境中,员工需应对更高的信息负荷与认知压力,其注意力资源主要集中于任务本身,难以充分感知与调节自身的情绪状态,从而削弱了人机信任对技术焦虑的缓解作用。相反,在任务复杂性较低时,员工拥有更多的认知资源去评估 AI 系统的可靠性,并更好地利用信任感带来的心理安全感,从而有效降低技术焦虑。此外,复杂任务通常赋予员工更高的自主性与效能感,也可能间接缓解其对技术系统的过度依赖与焦虑感^[29]。据此,提出以下假设

H₄:任务复杂性负向调节人机信任和技术焦虑之间的关系。对于任务复杂性较高的作业,人机信任对技术焦虑的负向影响更弱。

基于 H₁ 和 H₃,研究认为任务复杂性在风险感知中介路径中,可以起正向调节中介作用。对于任务复杂性水平较高的情况,对 AI 系统信任的矿工更容易削弱其风险感知能力,从而增加不安全行为的产生。结合 H₂ 和 H₄,研究认为任务复杂性在技术焦虑中介路径中,可以起负向调节中介作用。对于任务复杂性水平较低的情况,对 AI 系统信任的矿工更容

易减少技术焦虑,从而减少不安全行为。据此,提出以下假设

H₅:任务复杂性正向调节风险感知在人机信任和不安全行为之间的中介作用。

H₆:任务复杂性负向调节技术焦虑在人机信任和不安全行为之间的中介作用。

本文基于认知-情感个性系统理论和社会认知理论,系统分析人机信任通过风险感知和技术焦虑的中介作用对不安全行为的影响路径,并引入任务复杂性作为调节变量,探讨不同情境下人机信任对不安全行为的作用差异,研究的理论模型见图 1。研究假设的提出为后续实证研究提供了理论依据,同时为高危矿业环境中的信任管理与行为安全优化提供了理论支持。

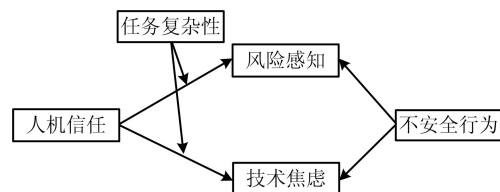


图 1 研究假设模型

Fig. 1 Research hypothesis model

2 研究方法

2.1 研究对象

本文研究采用分层随机抽样方法,在山西省、辽宁省、内蒙古自治区选取大中型煤矿企业各 15 家,以井下作业类型(采掘、通风、运输)为分层依据发放问卷。共发放问卷 600 份,收回 534 份。经筛选得到有效问卷 508 份,回收率为 84.7%。研究对象大多数为男性,占 96.4%;18~25 岁占比为 6.3%、26~35 岁占比为 43.5%、36~45 岁占比为 42.7%、46 岁及以上占比为 7.5%,其中 26~45 岁群体占比为 86.2%,为煤矿井下作业的主力年龄段;初中及以下学历占比为 19.3%、高中或中专占比为 41.5%、大专占比为 26.8%、本科及以上占比为 12.4%,其中大专及以下学历者合计占比为 87.3%;工龄≤1 a 占比为 4.7%、>1~2 a 占比为 8.9%、>2~5 a 占比为 59.8%、>5~10 a 占比为 19.3%、>10 a 占比为 7.3%,其中>2~5 a 工龄占比最大。样本构成基本符合我国煤矿从业人员的行业结构特征,样本代表性良好。

2.2 研究工具

研究选取国内外成熟量表测量主要变量以确保

问卷信效度,涉及的英文量表采取“翻译-回译”的方法进行转换。题项采用 likert-5 级量表法进行测量,1~5 分别代表被试对题项“非常不同意”到“非常同意”。

(1) 人机信任。根据 Gulati 等^[30]提出的人机信任量表,结合高危矿业场景进行改编,包括“该系统的行为让我怀疑其真实意图”等 12 个题项。采用 likert-5 点计分法进行研究,分数越高,人机信任水平越高。该量表 $\alpha = 0.842$ 。

(2) 风险感知。采用李广利^[31]提出的风险感知量表,通过风险知识及态度和风险后果认知两个维度进行研究,包括“工作中的风险是可以观测到的吗”“您对工作中遇到的风险熟悉吗”等 10 个题项。采用 likert-5 点计分法进行研究,分数越高,矿工的风险感知能力越强,该量表 $\alpha = 0.874$ 。

(3) 技术焦虑。采用 Wang 等^[32]开发的人工智能焦虑量表,选择“学习”与“工作替代”两个维度进行测量,包括“学会理解与技术/产品相关的所有特殊功能让我感到焦虑”等 6 个题项。采用 likert-5 点计分法进行研究,分数越高,矿工的技术焦虑程度越高,该量表 $\alpha = 0.945$ 。

(4) 不安全行为。采用王丹^[33]绘制的不安全行为量表,通过安全不遵守行为与安全不参与行为两个维度进行研究,包括“我有过不遵守操作规程或标准的行为”“我一般不与上级沟通”等 9 个题项。采用 likert-5 点计分法进行研究,分数越高,矿工不安全行为发生频次越高。该量表 $\alpha = 0.920$ 。

(5) 任务复杂性。根据 Kuhn 等^[34]的研究,结合

本文研究的特点进行适应性调整,包括“完成我与 AI 协同进行的任务需要付出很多精力”等 6 个题项。采用 likert-5 点计分法进行研究,分数越高,任务复杂性越大,该量表 $\alpha = 0.820$ 。

3 数据结果与分析

3.1 共同方法偏差

为检验所采集测量数据的共同方法偏差,研究通过匿名方式获取数据,以在一定程度上减少作答者对题项的猜想与戒备^[35]。首先,采用 Harman 单因素法进行分析,结果显示,第 1 个析出因子的因子载荷量为 30.189%,未超过 40%,且 5 个特征值大于 1,初步可以判定本文研究的共同方法偏差并不严重。其次,通过 AMOS 24.0 软件在五因子模型的基础之上加入 1 个共同方法因子。结果显示,模型的拟合度未得到明显改善 ($\chi^2/\nu = 1.362$, CFI 为 0.986, TLI 为 0.983, RMSEA 为 0.017, SRMR 为 0.025), ΔCFI 和 ΔTLI 未超过 0.1, $\Delta RMSEA$ 和 $\Delta SRMR$ 未超过 0.05,由此可以确定不存在严重的共同方法偏差问题。

3.2 描述性统计分析

主要变量的平均值、标准差及变量间相关系数见表 1。由表 1 可知:人机信任与风险感知 ($r = -0.343, p < 0.01$)、技术焦虑 ($r = -0.568, p < 0.01$) 及不安全行为 ($r = -0.391, p < 0.01$) 之间均存在显著相关关系。风险感知与不安全行为 ($r = -0.377, p < 0.01$)、技术焦虑与不安全行为 ($r = 0.682, p < 0.01$) 之间也显著相关。这些相关性与理

表 1 均值、标准差和相关系数

Table 1 Means, standard deviations, and correlation coefficients

变量	均值	标准差	相关系数								
			1	2	3	4	5	6	7	8	9
1. 性别	1.26	0.440	1								
2. 年龄	2.35	0.787	0.202 **	1							
3. 学历	1.77	0.854	0.233 **	0.014	1						
4. 工作年限	2.36	1.104	0.209 **	0.818 **	-0.002	1					
5. 人机信任	2.931	0.686	0.069	-0.011	0.017	0.031	1				
6. 风险感知	4.086	0.670	0.077	-0.006	0.081	-0.003	-0.343 **	1			
7. 技术焦虑	3.782	0.690	0.089 *	-0.002	0.050	0.001	-0.568 **	-0.345 **	1		
8. 不安全行为	3.576	0.675	0.108 *	-0.058	-0.028	-0.019	-0.391 **	-0.377 **	0.682 **	1	
9. 任务复杂性	3.482	0.780	0.049	-0.058	-0.072	0.032	0.303 *	0.659 **	0.692 **	0.576 **	1

注: * 代表 $p < 0.05$, ** 代表 $p < 0.01$ 。

论预期一致,因此,研究假设得到初步支持。

3.3 验证性因子分析

运用 AMOS 24.0 软件进行验证性因子分析,见表2。由表2可知:五因子模型在各模型中拟合最佳($\chi^2/\nu = 1.529$, CFI 为 0.952, TLI 为 0.958, RMSEA 为 0.047),各变量区分效度良好,可以进行后续检验。

3.4 假设检验

假设1认为人机信任通过削弱风险感知间接增加不安全行为。由表3可见,人机信任与风险感知显著负相关($b = -0.739, p < 0.001$),风险感知显著负向影响不安全行为($b = -0.471, p < 0.001$)。根据模型8可知,在加入风险感知变量后,人机信任显著影响不安全行为($b = -0.139, p < 0.001$)。此外,表4的中介效应分析结果显示,人机信任通过风险

感知影响不安全行为的间接效应为0.151,95%的置信区间为[0.087, 0.212],区间不包含0。因此,假设 H_1 得到验证。

假设2认为人机信任通过降低技术焦虑间接抑制不安全行为。由表3可见,人机信任与技术焦虑显著负相关($b = -0.542, p < 0.01$),技术焦虑显著正向预测不安全行为($b = 0.689, p < 0.001$)。根据模型10可知,人机信任与技术焦虑的交互项显著影响不安全行为($b = 0.552, p < 0.001$)。此外,表4的中介效应分析结果显示,人机信任通过技术焦虑影响不安全行为的间接效应为-0.299,95%的置信区间为[-0.357, -0.241],区间不包含0。因此,假设 H_2 得到验证。

假设3认为任务复杂性对人机信任与风险感知具有调节作用。如表5中的模型2,人机信任和任务

表2 验证性因子分析
Table 2 Confirmatory factor analysis

模型描述	χ^2	ν	χ^2/ν	CFI	TLI	RMSEA
五因子模型(A、B、C、D、E)	1 364.248	892	1.529	0.952	0.958	0.047
四因子模型(A+B、C、D、E)	2 019.816	896	2.254	0.863	0.860	0.071
三因子模型(A+B、C+D、E)	2 463.527	899	2.740	0.765	0.754	0.063
二因子模型(A+B、C+D+E)	3 156.219	901	3.503	0.703	0.710	0.092
单因子模型(A+B+C+D+E)	4 028.315	903	4.461	0.369	0.385	0.085
加入共同方法因子	1 237.843	851	1.455	0.955	0.951	0.034

表3 人机信任、风险感知、技术焦虑与不安全行为的回归模型分析

Table 3 Regression model analysis of human-machine trust, risk perception, technological anxiety, and unsafe behaviors

变量类型	变量	风险感知		技术焦虑		不安全行为						
		模型1	模型2	模型3	模型4	模型5	模型6	模型7	模型8	模型9	模型10	模型11
控制变量	性别	0.059	0.005	0.079	0.039	0.131 **	0.090 *	0.103 *	0.089	0.076	0.068	0.068
	年龄	-0.020	0.067	-0.020	0.045	-0.141	-0.075	-0.131	-0.085	-0.127	-0.100	-0.096
	学历	0.068	0.067 *	0.032	0.031	-0.056	-0.057	-0.088 *	-0.066	-0.078 *	-0.074 *	-0.069
	工龄	0.001	-0.082	0	-0.060	0.069	0.007	0.068	0.018	0.069	0.040	0.036
自变量	人机信任	-0.739 ***		-0.542 ***		-0.552 ***		-0.450 ***		-0.253 ***		
中介变量	风险感知							-0.471 ***		-0.139 *		
	技术焦虑									0.689 ***		
										0.552 ***		
	R^2	0.010	0.551	0.008	0.299	0.022	0.324	0.242	0.333	0.493	0.538	0.540
	ΔR^2	0.010	0.541	0.008	0.291	0.022	0.302	0.242	0.311	0.471	0.516	0.518
	F	1.250	123.027 ***	1.034	42.794 ***	2.822 *	48.116 ***	31.995 ***	41.614 ***	97.575 ***	97.124 ***	83.763 ***

注: R^2 为模型解释力度; ΔR^2 为模型变化时 R^2 的变化情况; F 用于判断模型是否有意义,若 $p < 0.05$,表明模型有意义;*表示 $p < 0.05$,**表示 $p < 0.01$,***表示 $p < 0.001$ 。

复杂性的交互项显著影响风险感知($b = 0.195, p < 0.01$)。为直观呈现这一调节效应,本文研究开展简单斜率分析,如图 2 所示,分别取高、低复杂性任务水平绘制交互效应图。结果显示,当任务复杂性较高时,人机信任对风险感知的负向影响更为显著,从而验证了假设 H_3 。

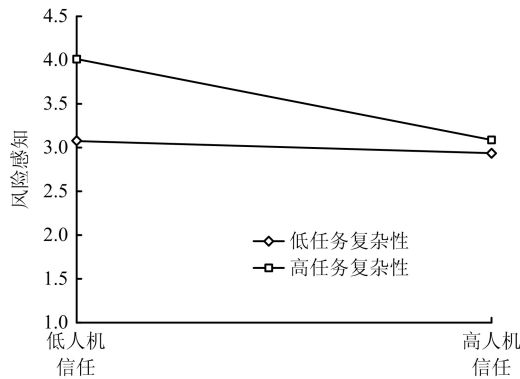


图 2 风险感知的调节效应

Fig. 2 Moderating effect of risk perception

假设 4 认为任务复杂性对人机信任与技术焦虑之间的关系具有调节作用,人机信任和任务复杂性的交互项显著影响技术焦虑($b = -0.183, p < 0.01$)。如图 3 所示,通过绘制不同任务复杂性水平的交互效应图发现,当任务复杂性较高时,人机信任对技术焦虑的负向影响更弱。由此,假设 H_4 得到验证。

假设 5 提出任务复杂性正向调节风险感知在人机信任与不安全行为之间的消极中介作用。如表 6 所示,风险感知的调节中介效应指数为 0.080,95% 置信区间为 $[0.035, 0.129]$ 。在任务复杂性较高时,风险感知的中介效应显著,效应值为 0.168,95% 置信区间为 $[0.118, 0.214]$;在任务复杂性较低时,风险感知的中介效应显著,效应值为 0.093,95% 置信区间为 $[0.045, 0.148]$,假设 H_5 得到验证。

假设 6 提出任务复杂性负向调节技术焦虑在人

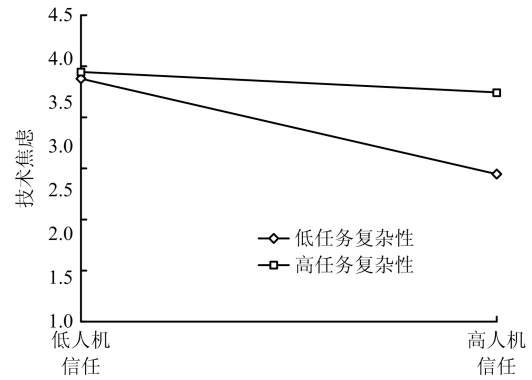


图 3 技术焦虑的调节效应

Fig. 3 Moderating effect of technological anxiety

机信任与不安全行为之间的积极中介作用。如表 6 所示,技术焦虑的调节中介效应指数为 0.057,95% 置信区间为 $[0.029, 0.087]$ 。在任务复杂性较高时,技术焦虑的间接效应显著,效应值为 -0.064 ,95% 置信区间为 $[-0.114, -0.016]$;在任务复杂性较低时,技术焦虑的间接效应显著,效应值为 -0.171 ,95% 置信区间为 $[-0.236, -0.117]$ 。假设 H_6 得到支持。

4 讨论

4.1 人机信任与不安全行为

人机信任对不安全行为的影响呈双刃剑效应。一方面,人机信任通过降低矿工风险感知水平,显著增加不安全行为发生率;另一方面,人机信任通过缓解技术焦虑间接抑制不安全行为。与以往研究强调的人机信任正向作用^[36]不同,本文研究揭示了人机信任在认知与情感路径中作用方向的不一致,突破了传统研究的单一视角,为智能矿山的安全管理提供了新的理论支持。煤矿企业在推动智能系统应用时,应重视适度信任。可通过透明化设计、作业风险警示等手段,避免矿工因过度信任而弱化主观判断。同时,引导员工在信任系统的同时,保留风险警觉与操作自律。

表 4 风险感知和技术焦虑的多重中介效应检验

Table 4 Tests of the multiple mediation effects of risk perception and technological anxiety

效应	路径	效应值	标准误差	95% 置信区间
间接效应 1	人机信任→风险感知→不安全行为	0.151	0.035	[0.087, 0.212]
间接效应 2	人机信任→技术焦虑→不安全行为	-0.299	0.030	[-0.357, -0.241]
总中介效应	间接效应 1 + 间接效应 2	-0.148	0.040	[-0.227, -0.068]
对比中介效应	间接效应 1 - 间接效应 2	0.450	0.051	[0.373, 0.526]

表5 任务复杂性对风险感知、技术焦虑的调节作用

Table 5 Moderating effects of task complexity on risk perception and technological anxiety

变量类型	变量	风险感知		技术焦虑	
		模型 1	模型 2	模型 3	模型 4
控制变量	性别	0.032	0.002	0.055	0.034
	年龄	0.048	0.076	0.042	0.060
	学历	0.101 *	0.073	0.061	0.042
	工龄	-0.049	-0.287	-0.044	-0.070
自变量	人机信任		-0.267 ***		-0.284 ***
调节变量	任务复杂性	0.380 ***	0.271 ***	0.339 ***	0.215 ***
	交互项		0.195 **		-0.183 **
	R^2	0.151	0.554	0.121	0.311
	ΔR^2	0.142	0.403	0.113	0.190
	F	17.910 ***	58.777 ***	13.786 ***	32.239 ***

注: R^2 为模型解释力度; ΔR^2 为模型变化时 R^2 的变化情况; F 用于判断模型是否有意义, 若 $p < 0.05$, 表明模型有意义; * 表示 $p < 0.05$, ** 表示 $p < 0.01$, *** 表示 $p < 0.001$ 。

表6 任务复杂性的调节中介效应

Table 6 Moderated mediation effects of task complexity

中介变量	指标	效应值	标准误差	95% 置信区间
风险感知	低水平	0.093	0.025	[0.045, 0.148]
	平均值	0.130	0.016	[0.082, 0.181]
	高水平	0.168	0.026	[0.118, 0.214]
	有调节的中介效应	0.080	0.013	[0.035, 0.129]
技术焦虑	低水平	-0.171	0.028	[-0.236, -0.117]
	平均值	-0.117	0.021	[-0.165, -0.067]
	高水平	-0.064	0.029	[-0.114, -0.016]
	有调节的中介效应	0.057	0.015	[0.029, 0.087]

4.2 风险感知的中介效应

风险感知在人机信任与不安全行为之间起中介作用。在矿工与 AI 系统交互中, 当各项因素导致矿工对 AI 系统不信任时, 矿工会高估风险, 即风险感知能力增强, 从而采取更谨慎的行为, 减少不安全行为; 而当矿工对 AI 系统信任时, 可能会低估作业环境中的潜在风险, 放松对自身行为的约束, 风险感知能力减弱, 进而提高不安全行为的发生概率。这与 Kenesei 等^[37]在自动驾驶领域发现的信任影响风险感知, 进而产生不同的行为决策保持一致。在实际作业管理中, 煤矿企业应通过定期组织风险认知培训、模拟演练等方式, 帮助矿工维持对风险的合理敏

感度。此外, 设计 AI 系统时可适当引入风险提示与反馈模块, 促使矿工在信任系统的同时, 保持应有的风险警惕性。

4.3 技术焦虑的中介效应

技术焦虑在人机信任与不安全行为之间起中介作用。在矿工与 AI 系统的交互过程中, 当对 AI 系统不信任时, 会对系统的可靠性和安全性产生怀疑, 进而增加对技术的焦虑感, 导致矿工的心理压力增加, 影响其工作中的情绪状态和行为表现, 引发安全事故的发生。当矿工对 AI 系统信任程度较高时, 技术焦虑则会得到缓解。这种缓解的技术焦虑有助于矿工保持积极的情绪状态, 更好地执行安全操作规

程,从而减少不安全行为的发生。与仅从技术适应视角理解技术焦虑的研究不同,本文研究强调了信任对情绪状态的缓冲作用。在矿山智能化转型过程中,需关注一线作业人员的情绪状态。煤矿企业可通过开展心理减压活动,缓解技术引入带来的不适与压力,提升作业情绪稳定性,降低不安全行为发生概率。

4.4 任务复杂性的调节效应

任务复杂性在人机信任对风险感知的影响中起正向调节作用,当任务复杂性水平较高时,矿工对 AI 系统的信任可能会使其过度依赖其决策和建议,忽视自身对风险的判断和经验,从而进一步削弱风险感知能力,更易诱发不安全行为的发生;而任务复杂性在人机信任对技术焦虑的影响中起负向调节作用,其复杂性越高,矿工对 AI 系统的信任缓解焦虑的效果越弱。这表明在复杂作业环境下,矿工的心理资源会更更多地被任务本身所占用,难以从信任 AI 系统中获得足够的心理支持来缓解技术焦虑。相比于只关注任务特征对操作绩效的研究^[38],本文研究揭示了任务复杂性如何干预信任机制效应路径,为理解复杂环境下的人机交互模式提供了新视角。在高复杂任务场景中,煤矿企业管理者应合理调配人机职责分工,避免操作人员过度依赖系统导致的判断懈怠。同时,应通过语音提示或视觉辅助等方式,降低认知负荷,提升矿工在复杂任务中的风险识别能力与情绪调节能力。

5 结 论

(1) 人机信任对不安全行为具有显著的“双刃剑”效应。具体而言,人机信任能够通过降低矿工的风险感知进而增加不安全行为的发生概率,但同时也能通过缓解技术焦虑,间接抑制不安全行为。该发现揭示了人机信任的复杂影响,一方面有助于煤矿企业提升采纳安全技术的主动性,另一方面也提示需关注因信任失当导致的操作失误风险。

(2) 任务复杂性在人机信任与不安全行为的影响路径中发挥了重要调节作用。研究结果显示,随着任务复杂性的提升,人机信任对风险感知的降低效应进一步增强,从而更易诱发不安全行为;而在技术焦虑路径中,任务复杂性则减弱了人机信任缓解技术焦虑的积极影响。这表明在复杂作业环境下,如何平衡人机信任成为防范不安全行为的关键。

(3) 研究数据主要来源于矿工自评问卷,存在共同方法偏差的可能,未来可结合行为观测和试验数据验证结果的稳健性。此外,研究情境相对单一,未来可拓展至其他高危行业智能化场景,由于篇幅原因,本文只考虑风险感知和技术焦虑两个中介因素以及任务复杂性一个调节变量,未来可以探讨认知路径中的安全效能感与感知行为控制,情感路径中的心理安全感与安全疲劳多重中介因素以及系统透明度等多重调节因素,丰富对人机信任效应边界的认知。

参考文献 (References):

- [1] 王启飞, 王俊龙, 刘昊霖, 等. 煤矿安全风险分析的文本数据模型与集成分析平台[J]. 安全与环境学报, 2024, 24(11): 4358 - 4365.
WANG Q F, WANG J L, LIU H L, et al. Text data model and integrated analysis platform for coal mine safety risk analysis[J]. Journal of Safety and Environment, 2024, 24(11): 4358 - 4365.
- [2] 牛莉霞, 李肖萌. 5G 时代面向情报需求的智慧矿山应急管理新模式研究[J]. 灾害学, 2024, 39(3): 228 - 234.
NIU L X, LI X M. Research on the new model of intelligent mine emergency management oriented to intelligence needs in the 5G era[J]. Journal of Catastrophology, 2024, 39(3): 228 - 234.
- [3] 袁晓芳, 伍琦淇, 孙林辉, 等. 智能煤矿下操作人员情境意识影响因素的多面分析[J]. 中国安全生产科学技术, 2024, 20(10): 118 - 124.
YUAN X F, WU Q Q, SUN L H, et al. Multifaceted analysis on influencing factors of operators' situation awareness in intelligent coal mine[J]. Journal of Safety and Technology, 2024, 20(10): 118 - 124.
- [4] YANG J, YANG X, CHAI S, et al. Research on the mechanism of human-machine security collaboration of miners considering automation trust[J]. Heliyon, 2024, 10(23): 367 - 394.
- [5] 田水承, 冯帅, 高晓颖, 等. 一线矿工不安全状态的熵权 TOPSIS 法评价[J]. 安全与环境学报, 2023, 23(11): 4049 - 4056.
TIAN S C, FENG S, GAO X Y, et al. Study on the evaluation of entropy-weighted TOPSIS method for the unsafe state of tier miners[J]. Journal of Safety and Environment, 2023, 23(11): 4049 - 4056.
- [6] 孙剑, 黄倩倩, 章伯阳, 等. 事故体验对建筑工人不

- 安全行为意向的影响机理研究[J]. 安全与环境学报, 2025, 25(4): 1424–1433.
- SUN J, HUANG Q Q, ZHANG B Y, et al. Investigating the influence mechanism of accident experience on construction workers' intentions to engage in unsafe behaviors[J]. Journal of Safety and Environment, 2025, 25(4): 1424–1433.
- [7] 黄玥诚, 李泊宁, 郁晓霞, 等. 依恋关系在施工团队成员安全互动中的作用路径[J]. 清华大学学报(自然科学版), 2023, 63(2): 169–178.
- HUANG Y C, LI B N, YU X X, et al. Path analysis of attachment in the safety interactions of construction team members[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(2): 169–178.
- [8] MUIR B M. Trust between humans and machines, and the design of decision aids[J]. International Journal of Man-Machine Studies, 1987, 27(5/6): 527–539.
- [9] WICZOREK R, MANZEY D. The dark side of trust: the impact of automation misuse on human performance in safety-critical environments[J]. Safety Science, 2019, 115: 27–37.
- [10] HANCOCK P A, BILLINGS D R, SCHAEFER K E, et al. A meta-analysis of factors affecting trust in human-robot interaction[J]. Human Factors, 2011, 53(5): 517–527.
- [11] CHIEN S, DING J, CHUANG C, et al. Effects of trust and cognitive load on interaction with intelligent systems in automated driving[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2020, 74: 1–14.
- [12] WINTERSBERGER P, NOAH B E, KRAUS J, et al. Second workshop on trust in the age of automated driving [C]//Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, September 23, 2018, Toronto, Canada. New York: Association for Computing Machinery, 2018.
- [13] 胡飞翔, 周建亮, 林欣燕, 等. 情绪对建筑工人不安全行为影响的认知试验[J]. 中国安全科学学报, 2023, 33(4): 75–83.
- HU F X, ZHOU J L, LIN X Y, et al. A cognitive experiment of effect of emotion on construction workers' unsafe behavior[J]. China Safety Science Journal, 2023, 33(4): 75–83.
- [14] 李乃文, 武兴波, 牛莉霞. 非正式互动对新生代矿工不安全行为的影响[J]. 安全与环境学报, 2024, 24(6): 2329–2335.
- LI N W, WU X B, NIU L X. Impact of informal interaction on the unsafe behavior of new generation miners[J]. Journal of Safety and Environment, 2024, 24(6): 2329–2335.
- [15] 齐玥, 陈俊廷, 秦邵天, 等. 通用人工智能时代的人与AI信任[J]. 心理科学进展, 2024, 32(12): 2124–2136.
- QI Y, CHEN J T, QIN S T, et al. Human-AI mutual trust in the era of artificial general intelligence[J]. Advances in Psychological Science, 2024, 32(12): 2124–2136.
- [16] BANDURA A. Human agency in social cognitive theory[J]. American Psychologist, 1989, 44(9): 1175–1184.
- [17] 石娟, 李宵香, 郑光宇, 等. 社会网络视角下建筑工人不安全行为传播仿真研究[J]. 安全与环境学报, 2022, 22(2): 859–867.
- SHI J, LI X X, ZHENG G Y, et al. Study on propagation simulation of unsafe behaviors of construction workers from the perspective of social network[J]. Journal of Safety and Environment, 2022, 22(2): 859–867.
- [18] 郭衍宏, 魏旭. 老龄员工技术焦虑对人工智能技术持续采纳的影响: 组织信任与ICT自我效能的遮掩效应[J]. 软科学, 2024, 38(6): 7–12.
- GUO Y H, WEI X. Aging employees' technology anxiety and continuous adoption of artificial intelligence technology: the masking effect of organizational trust and ICT self-efficacy[J]. Soft Science, 2024, 38(6): 7–12.
- [19] SLOVIC P. Perception of risk[J]. Science, 1987, 236(4799): 280–285.
- [20] 朱永跃, 余莉花. 数字化领导力对制造业员工敏捷性的影响: 基于认知-情感双路径视角[J]. 科技进步与对策, 2024, 41(24): 140–149.
- ZHU Y Y, YU L H. The impact of digital leadership on employees' agility in manufacturing industry: a dual-path perspective of cognition and emotion[J]. Science & Technology Progress and Policy, 2024, 41(24): 140–149.
- [21] 陈登航, 王硕, 汪琛. 社会信任系统何以降低公众的人工智能风险感知?: 基于2023年科技与社会晴雨表调查的实证分析[J/OL]. 科学学研究: 1–16 [2025-04-12]. <https://doi.org/10.16192/j.cnki.1003-2053.20250313.003>.
- CHEN D H, WANG S, WANG C. How can social trust

- systems reduce public's AI risk perception? Empirical analysis based on the 2023 science, technology and society barometer survey[J/OL]. *Studies in Science of Science*: 1 – 16 [2025 – 04 – 12]. <https://doi.org/10.16192/j.cnki.1003-2053.20250313.003>.
- [22] 李翊均, 杨珂, 罗讯, 等. 隧道工人认知偏差对其不安全行为的影响研究: 风险感知的中介效应[J]. *中国安全科学学报*, 2024, 34(10): 24 – 29.
- LI H J, YANG K, LUO X, et al. Effects of tunnel workers' cognitive bias on their unsafe behaviors: mediating role of risk perception [J]. *China Safety Science Journal*, 2024, 34(10): 24 – 29.
- [23] 赵磊磊, 陈祥梅, 马志强. 人工智能时代教师技术焦虑: 成因分析与消解路向[J]. *首都师范大学学报(社会科学版)*, 2022(6): 138 – 149.
- ZHAO L L, CHEN X M, MA Z Q. Teachers' technology anxiety in the era of artificial intelligence: causes and approaches to alleviation[J]. *Journal of Capital Normal University (Social Sciences Edition)*, 2022 (6): 138 – 149.
- [24] CHOUNG H, DAVID P, ROSS A. Trust in AI and its role in the acceptance of AI technologies [J]. *International Journal of Human-Computer Interaction*, 2023, 39(9): 1727 – 1739.
- [25] YANG X, TIAN Y, FENG K, et al. Signal game analysis on the effectiveness of coal mine safety supervision based on the affective events theory [J]. *Complexity*, 2020, 2020(1): 5710419.
- [26] XIONG Y, SHI Y, PU Q, et al. More trust or more risk? User acceptance of artificial intelligence virtual assistant [J]. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 2024, 34 (3): 190 – 205.
- [27] HOPKO S K, KHURANA R, MEHTA R K, et al. Effect of cognitive fatigue, operator sex, and robot assistance on task performance metrics, workload, and situation awareness in human-robot collaboration [J]. *IEEE Robotics and Automation Letters*, 2021, 6(2): 3049 – 3056.
- [28] 江静, 董雅楠, 武欣, 等. 领导批判性思维能激活团队创造力吗? 一个三重交互作用模型[J]. *管理评论*, 2024, 36(8): 185 – 199.
- JIANG J, DONG Y N, WU X, et al. Will leader critical thinking trigger team creativity? A three-way interactive model [J]. *Management Review*, 2024, 36 (8): 185 – 199.
- [29] HUMPHREY S E, NAHRGANG J D, MORGESON F P. Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature [J]. *Journal of Applied Psychology*, 2007, 92(5): 1332 – 1356.
- [30] GULATI S, SOUSA S, LAMAS D. Design, development and evaluation of a human-computer trust scale [J]. *Behaviour & Information Technology*, 2019, 38(10): 1004 – 1015.
- [31] 李广利. 煤矿员工风险感知量表开发与应用[J]. *中国安全生产科学技术*, 2022, 18(7): 128 – 134.
- LI G L. Development and application of risk perception scale for coal mine employees [J]. *Journal of Safety Science and Technology*, 2022, 18(7): 128 – 134.
- [32] WANG Y Y, WANG Y S. Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior [J]. *Interactive Learning Environments*, 2022, 30(4): 619 – 634.
- [33] 王丹. 辱虐管理对矿工不安全行为的影响研究[J]. *经济与管理*, 2012, 26(10): 65 – 70.
- WANG D. Abusive supervision on the miners' unsafe behavior influence research [J]. *Economy and Management*, 2012, 26(10): 65 – 70.
- [34] KUHN T, POOLE M S. Do conflict management styles affect group decision making? Evidence from a longitudinal field study [J]. *Human Communication Research*, 2000, 26(4): 558 – 590.
- [35] 牛莉霞, 杜瑞欣, 夏文德. 作业场所安全标志、风险感知与安全注意力关系[J]. *安全与环境学报*, 2024, 24(4): 1495 – 1504.
- NIU L X, DU R X, XIA W D. Relationship between workplace safety signs, risk perception and safety attention[J]. *Journal of Safety and Environment*, 2024, 24(4): 1495 – 1504.
- [36] MADHAVAN P, WIEGMANN D A. Similarities and differences between human-human and human-automation trust: an integrative review [J]. *Theoretical Issues in Ergonomics Science*, 2007, 8(4): 277 – 301.
- [37] KENESEI Z, ASVANYI K, KOKENY L, et al. Trust and perceived risk: how different manifestations affect the adoption of autonomous vehicles[J]. *Transportation Research Part A: Policy and Practice*, 2022, 164: 379 – 393.
- [38] ENDSLEY M R. Toward a theory of situation awareness in dynamic systems[M]//ENDSLEY M R, GARLAND D J. *Situational awareness*. London: Routledge, 2017: 9 – 42.

Exploring the influence mechanism of human-machine trust on miners' unsafe behaviors: a cognitive-emotional dual-path approach

NIU Lixia¹, ZHAO Yue¹, LI Guo², LI Xiaomeng¹

(1 College of Business Administration, Liaoning Technical University, Huludao 125105, Liaoning, China;

2 Zhongkuang Testing (Liaoning) Co., Ltd., Fuxin 123000, Liaoning, China)

Abstract: This study aims to elucidate the dual-path mechanisms through which human-machine trust influences unsafe behaviors among miners in intelligent coal mines. Specifically, it examines the mediating roles of risk perception and technological anxiety, along with the moderating role of task complexity, thereby addressing the double-edged nature of trust in human-machine collaboration. Grounded in the Cognitive-Affective Personality System theory and Social Cognitive theory, a dual-path impact model was developed. Data were collected via a stratified random survey conducted across 45 medium and large coal mining enterprises in Shanxi, Liaoning, and Inner Mongolia. A total of 600 questionnaires were distributed, yielding 508 valid responses (an effective rate of 84.7%). Reliability and validity were confirmed through Cronbach's α and confirmatory factor analysis. Hypotheses were tested using structural equation modeling and hierarchical regression analysis. The findings revealed that human-machine trust has a significant "double-edged sword" effect. On the cognitive pathway, it reduced miners' risk perception, which subsequently increased unsafe behaviors. Conversely, on the emotional pathway, it alleviated technological anxiety, thereby suppressing unsafe behaviors. Moderation tests indicated that task complexity amplified the negative influence of trust on risk perception, thereby strengthening the risk-increasing pathway. Conversely, it weakened the mitigating effect of trust on technological anxiety, attenuating the risk-reducing pathway. Moderated mediation analysis further confirmed that the indirect effects of risk perception and technological anxiety were contingent upon task complexity, demonstrating opposite adjustment trends. Overall, human-machine trust influences miners' unsafe behaviors through distinct cognitive and emotional mechanisms, with task complexity playing a significant role in shaping these pathways. The findings enhance the theoretical understanding of trust in high-risk intelligent work environments and underscore the necessity for calibrated trust management. From a practical perspective, coal enterprises should focus on improving system transparency, implementing risk-alert feedback mechanisms, and providing psychological support to achieve a balanced level of trust among miners. In complex tasks, managers should optimize human-machine role allocation and minimize cognitive load to ensure adequate risk awareness and emotional stability.

Key words: safe social sciences; human-machine trust; miners' unsafe behaviors; cognitive-emotional dual-path; risk perception; technological anxiety; task complexity